

Supplementary Materials

Materials & Methods

Various sources were used to complement the phylogenetic tree obtained from [1]: [2, 3, 4, 5, 6, 7].

References

- [1] Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D. and Peterson, K. J. (2011) The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals. *Science*, **334** (6059), 1091–1097.
- [2] Benton, M. J. and Ayala, F. J. (2003) Dating the Tree of Life. *Science*, **300** (5626), 1698–1700.
- [3] Gaunt, M. W. and Miles, M. A. (2002) An Insect Molecular Clock Dates the Origin of the Insects and Accords with Palaeontological and Biogeographic Landmarks. *Molecular Biology and Evolution*, **19** (5), 748–761.
- [4] Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A., Davis, M. P., Wainwright, P. C., Friedman, M. and Smith, W. L. (2012) Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences*, **109** (34), 13698–13703.
- [5] Peterson, K. J., Cotton, J. A., Gehling, J. G. and Pisani, D. (2008) The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363** (1496), 1435–1443.
- [6] Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A. *et al.* (2003) The Genome Sequence of *Caenorhabditis briggsae*: A Platform for Comparative Genomics. *PLoS Biology*, **1** (2).
- [7] Wiegmann, B. M., Trautwein, M. D., Kim, J.-W., Cassel, B. K., Bertone, M. A., Winterton, S. L. and Yeates, D. K. (2009) Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biology*, **7** (1), 34.

Results

Table S 1: Number of DACs with significantly enriched GO terms per family. Total number of DACs shown in parenthesis. Only DACs containing at least 3 proteins were analysed here. None of the DACs of the p53 family were large enough to be considered for this analysis.

Species	Nuclear Receptor	bZIP	Homeobox	C2H2 zinc finger	bHLH	p53
<i>Homo sapiens</i>	1 (2)	5 (6)	13 (18)	13 (57)	4 (5)	—
<i>Caenorhabditis elegans</i>	1 (2)	0 (3)	0 (6)	0 (7)	0 (1)	—
<i>Drosophila melanogaster</i>	0 (2)	0 (3)	1 (8)	1 (12)	1 (4)	—
<i>Danio rerio</i>	0 (2)	1 (7)	4 (19)	2 (76)	2 (6)	—

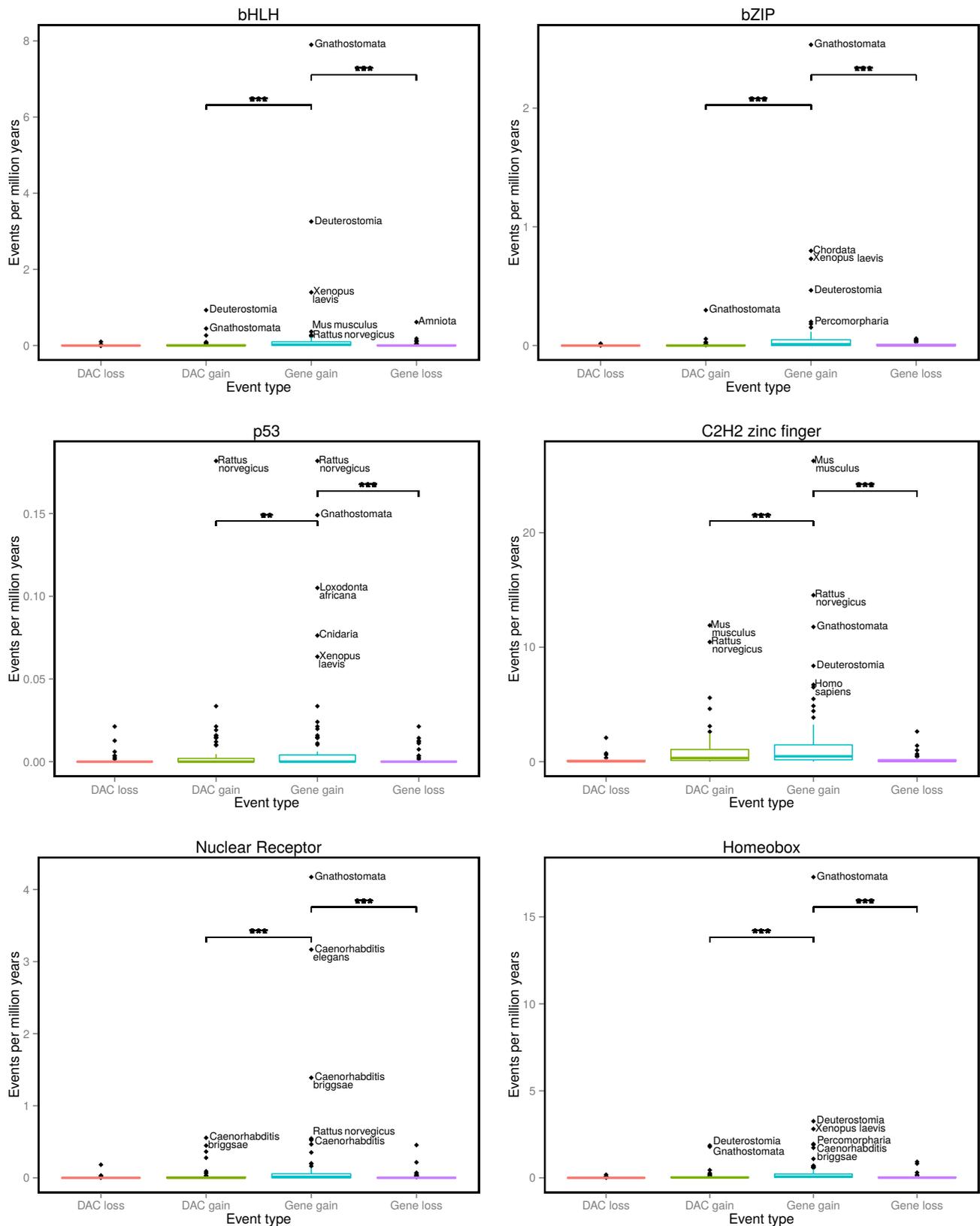


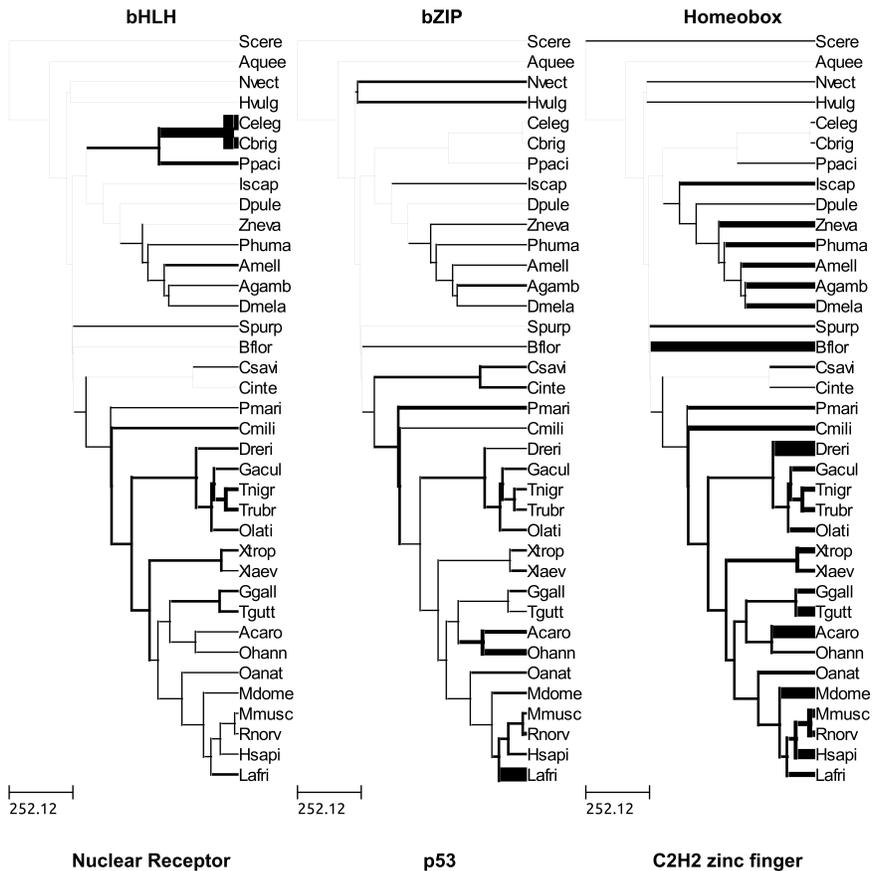
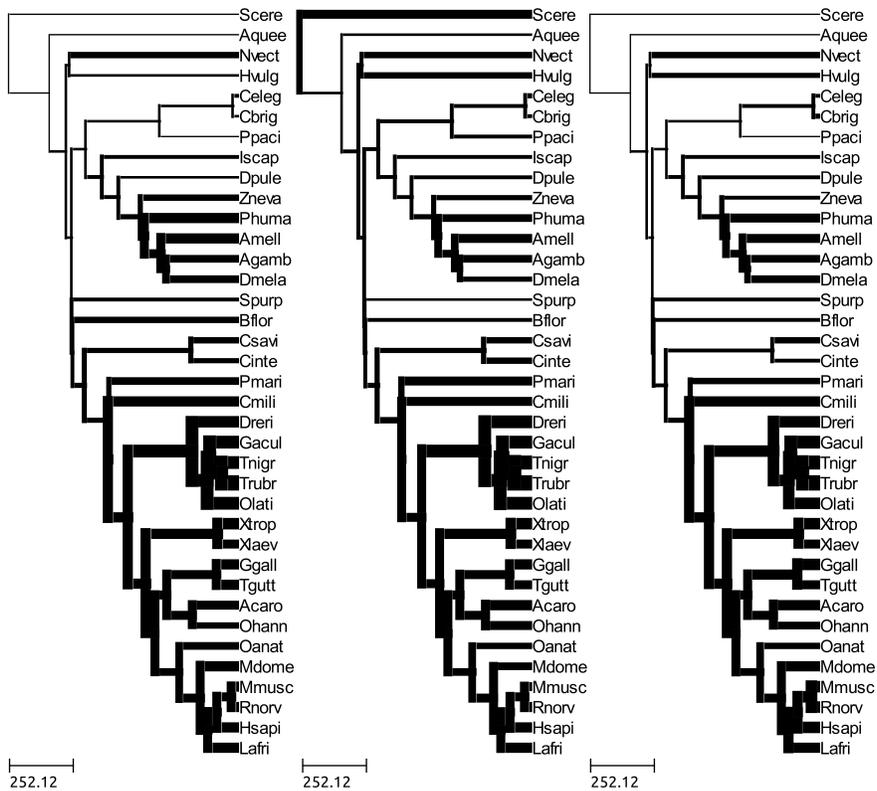
Figure S 1: Gene and DAC gain and loss rate distribution for all branches of the phylogenetic tree. Distribution of rates are shown as box plots, with one subplot used for each of the TF families. Significant differences between distributions of event types are marked with stars representing significance levels (**: $p < 0.01$, ***: $p < 0.001$). Branches are named after the node at the younger end of the branch.

Table S 2: Pfam HMMs assigned as TF family's DBDs. These HMMs were used to find the TF family members.

TF family	Pfam HMMs
bZIP	bZIP_1, bZIP_2, bZIP_Maf
p53	P53, CEP1-DNA_bind
homeobox	Homeobox, Homeobox_KN, Homez, CENP-B_N
NuclearFactor	zf-C4
bHLH	HLH
zfc2h2	zf-C2H2, zf-Di19, GAGA, zf-C2H2_jaz, zf-C2H2_2, zf-met, zf-H2C2_2, zf-C2H2_4, zf-H2C2_5, zf-C2H2_6, zf-C2HC_2

Table S 3: List of species used in this study with data sources. The Ensembl and Ensembl Metazoa databases can be found at <http://www.ensembl.org/> and <http://metazoa.ensembl.org/>, respectively.

Scientific name	NCBI TaxID	Common name	Data source
<i>Xenopus laevis</i>	8355	African clawed frog	JGI version 6.0 http://www.xenbase.org/genomes/static/laevis.jsp
<i>Danio rerio</i>	7955	Zebrafish	Ensembl
<i>Branchiostoma floridae</i>	7739	Amphioxus	JGI 1.0 http://genome.jgi-psf.org/Brafl1/Brafl1.home.html
<i>Callorhynchus milii</i>	7868	Elephant Shark	genome website, version from 2014/05/12 http://esharkgenome.imcb.a-star.edu.sg/
<i>Amphimedon queenslandica</i>	400682	Sponge	Ensembl Metazoa
<i>Nematostella vectensis</i>	45351	Starlet sea anemone	Ensembl Metazoa
<i>Caenorhabditis briggsae</i>	6238	Roundworm	Ensembl Metazoa
<i>Saccharomyces cerevisiae</i>	4932	Baker's yeast	Ensembl
<i>Ophiophagus hannah</i>	8665	King cobra	NCBI, from 2014/01/14
<i>Hydra vulgaris</i>	6087	Hydra	NCBI, from 2014/01/14
<i>Drosophila melanogaster</i>	7227	Fruit fly	Ensembl Metazoa
<i>Xenopus tropicalis</i>	8364	Western clawed frog	Ensembl
<i>Petromyzon marinus</i>	7757	Lamprey	Ensembl
<i>Daphnia pulex</i>	6669	Water flea	Ensembl Metazoa
<i>Anolis carolinensis</i>	28377	Green anole (lizard)	Ensembl
<i>Gasterosteus aculeatus</i>	69293	Stickleback	Ensembl
<i>Homo sapiens</i>	9606	Human	Ensembl
<i>Loxodonta africana</i>	9785	African bush elephant	Ensembl
<i>Caenorhabditis elegans</i>	6239	Roundworm	Ensembl Metazoa
<i>Mus musculus</i>	10090	Mouse	Ensembl
<i>Gallus gallus</i>	9031	Chicken	Ensembl
<i>Apis mellifera</i>	7460	European honey bee	Ensembl Metazoa
<i>Ciona savignyi</i>	51511	Sea squirt	Ensembl
<i>Rattus norvegicus</i>	10116	Common rat	Ensembl
<i>Tetraodon nigroviridis</i>	99883	Pufferfish	Ensembl
<i>Oryzias latipes</i>	8090	Japanese rice fish	Ensembl
<i>Pediculus humanus</i>	121225	Human body louse	Ensembl Metazoa
<i>Strongylocentrotus purpuratus</i>	7668	Purple sea urchin	Ensembl Metazoa
<i>Monodelphis domestica</i>	13616	Opossum	Ensembl
<i>Ixodes scapularis</i>	6945	Tick	Ensembl Metazoa
<i>Takifugu rubripes</i>	31033	Pufferfish	Ensembl
<i>Ornithorhynchus anatinus</i>	9258	Platypus	Ensembl
<i>Taeniopygia guttata</i>	59729	Zebra finch	Ensembl
<i>Anopheles gambiae</i>	7165	Mosquito	Ensembl Metazoa
<i>Ciona intestinalis</i>	7719	Sea squirt	Ensembl
<i>Pristionchus pacificus</i>	54126	Roundworm	Ensembl Metazoa
<i>Zootermopsis nevadensis</i>	136037	Termite	personal communication, version 2.2



Nuclear Receptor

p53

C2H2 zinc finger

Figure S 2: Phylogenetic tree showing measured and reconstructed TF family sizes in relationship to measured and reconstructed proteome sizes. The thickness of each branch is scaled according to the relationship between TF family and proteome size. Tree leaves are labeled according to the species names shortened as first letter of genus name + first four letters of species name.

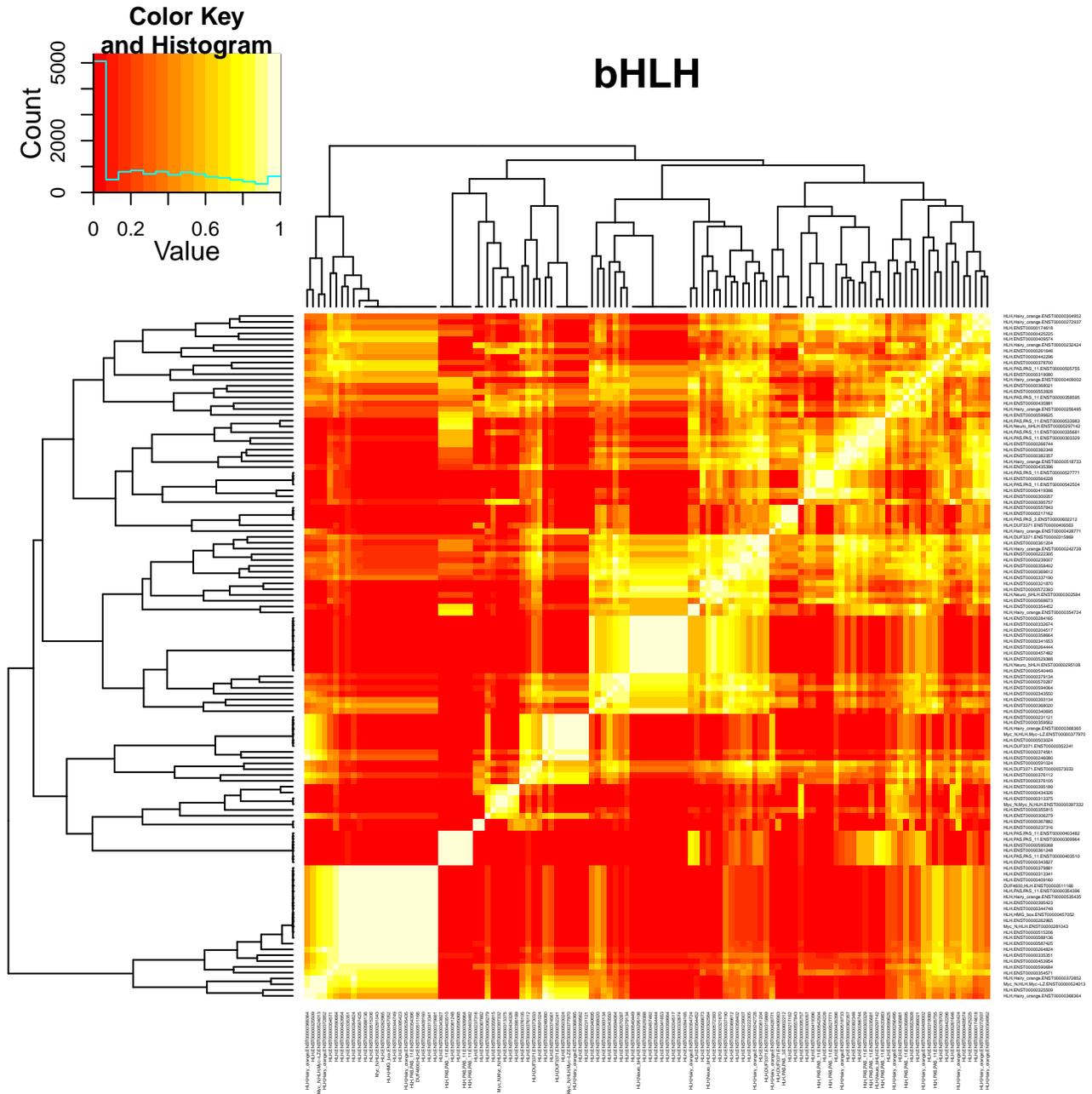


Figure S 3: Clustering of human TF family genes based on expression profiles. Heatmaps of cosine values are shown with gene names and domain arrangements as column and row labels. Online PDF version allows for zooming to read small labels.

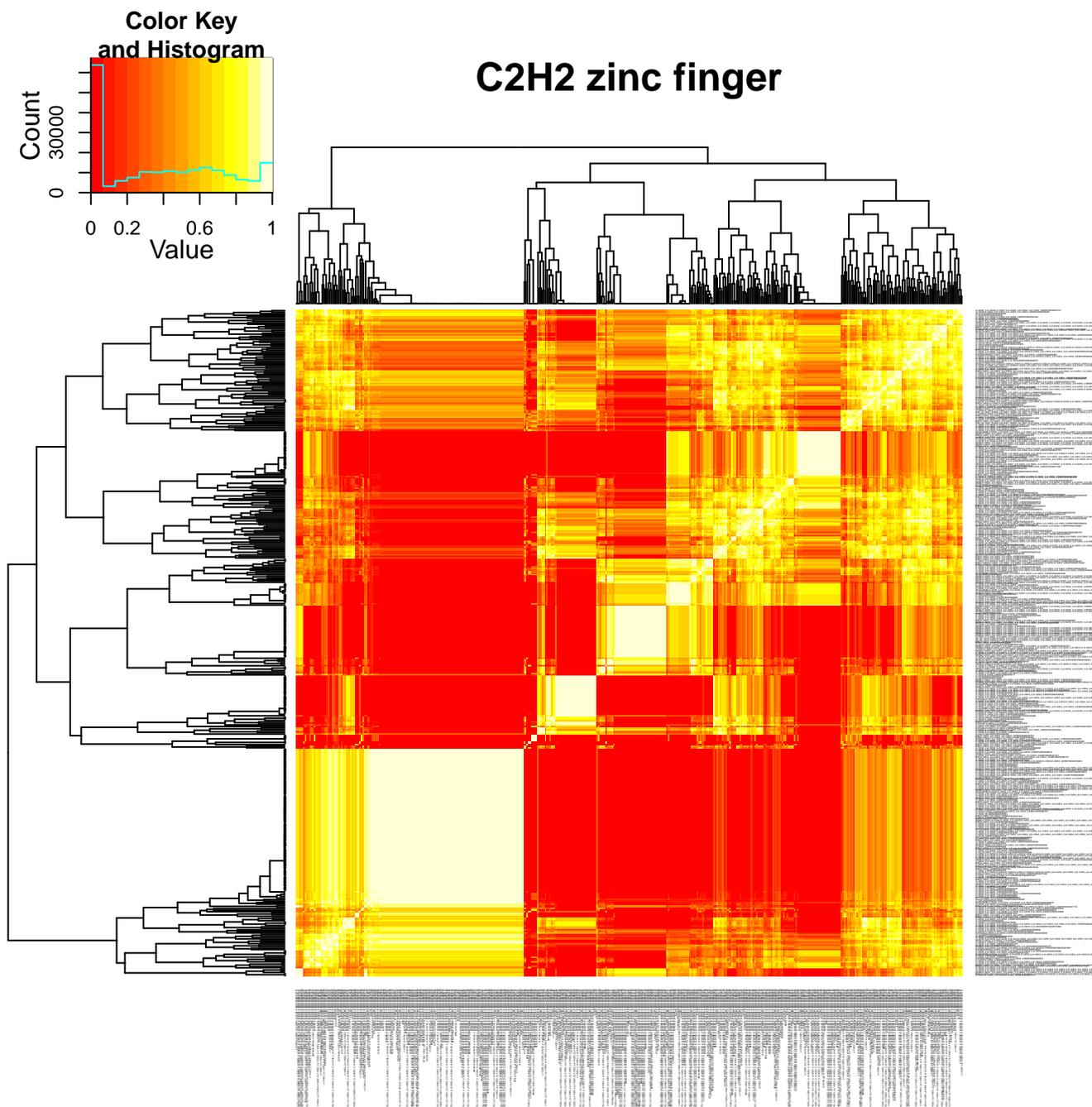


Figure S 3: *continued from previous page*

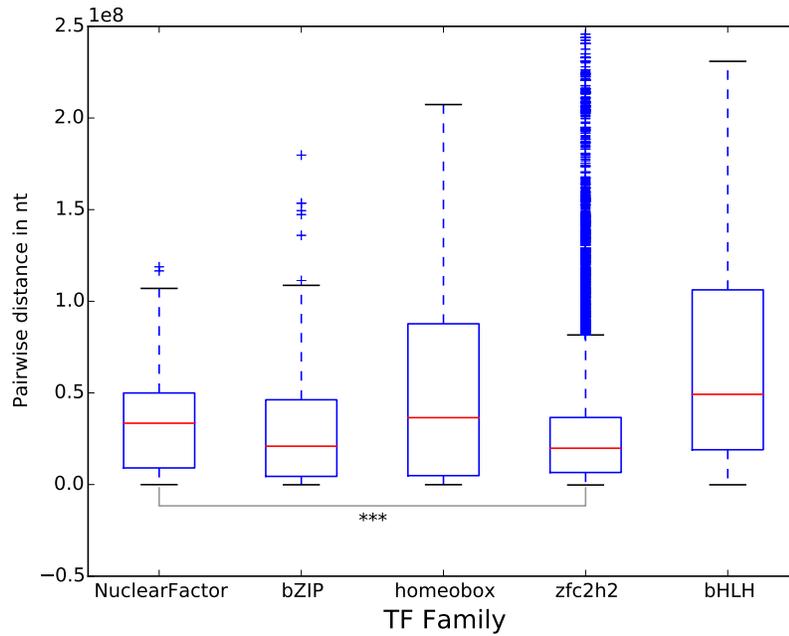


Figure S 4: Pairwise gene distances between genes found on the same chromosome in human. “***” denotes significant differences between distribution at level $p < 0.001$ (two-sided Kolmogorov-Smirnov test).

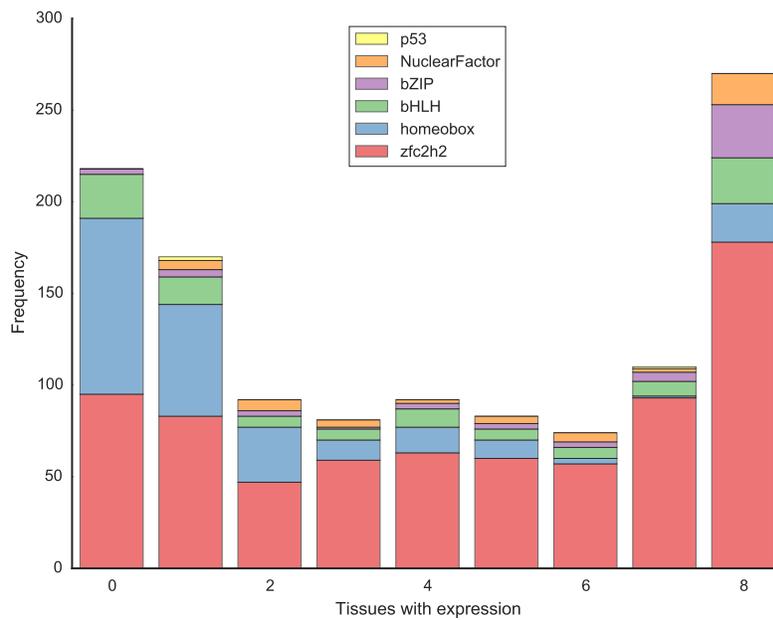


Figure S 5: Expression breadth of TFs of the analysed TF families. For each of the genes belonging to a TF family it was tested whether it was expressed in each of 8 organs (FPKM threshold 1).

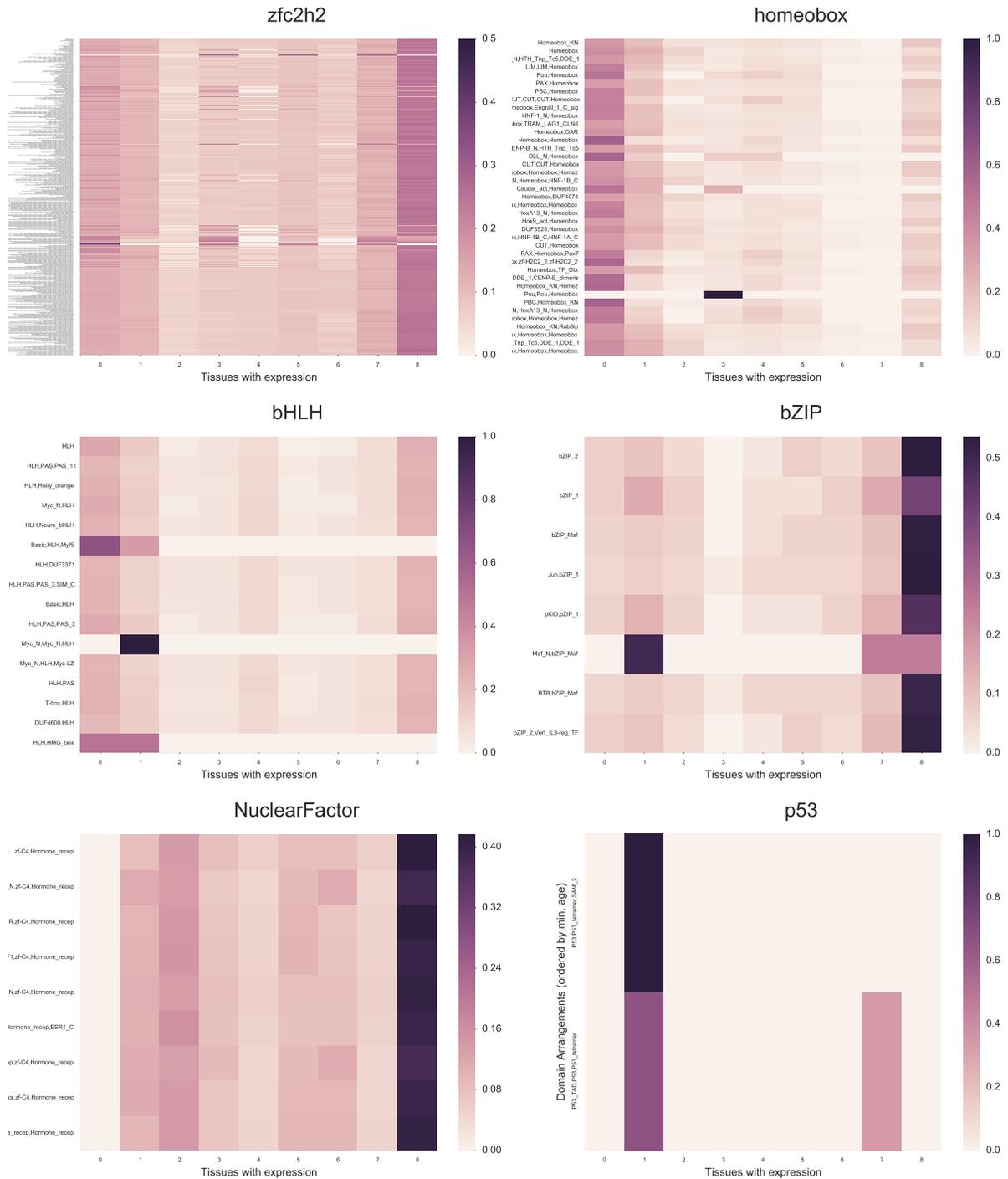


Figure S 6: Expression breadth of DACs of the analysed TF families. Each cell of the heatmap shows the relative fraction of genes of the depicted DAC expressed at the shown expression breadth. Darker colours stand for a higher fraction of genes being expressed in the respective amount of organs. DACs are ordered by age, older DACs are shown at the top, younger DACs at the bottom of heatmaps. Online PDF version allows for zooming to read small labels (zfc2h2 family).

Table S 4: GO terms found enriched in clusters of genes with similar expression of the bHLH family

Cluster	GO term	Term description	Genes with term in cluster	Genes with term in family	Fisher p-value
1	GO:0048731	system development	3/4	21/86	0.0435
5	GO:0060485	mesenchyme development	3/3	4/86	3.91e-05
5	GO:0048646	anatomical structure formation involved in morphogenesis	3/3	13/86	0.00279
5	GO:0009887	organ morphogenesis	3/3	13/86	0.00279
5	GO:0007389	pattern specification process	3/3	13/86	0.00279
5	GO:0009888	tissue development	3/3	16/86	0.00547
5	GO:0009653	anatomical structure morphogenesis	3/3	24/86	0.0198
5	GO:0048513	organ development	3/3	28/86	0.032
6	GO:0044444	cytoplasmic part	3/3	10/86	0.00117
6	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	3/3	30/86	0.0397
6	GO:0050896	response to stimulus	3/3	30/86	0.0397
7	GO:0065009	regulation of molecular function	4/4	19/86	0.00183
7	GO:0009968	negative regulation of signal transduction	3/4	9/86	0.00311
7	GO:0048585	negative regulation of response to stimulus	3/4	9/86	0.00311
7	GO:0010648	negative regulation of cell communication	3/4	10/86	0.00439
7	GO:0007219	Notch signaling pathway	3/4	10/86	0.00439
7	GO:0023057	negative regulation of signaling	3/4	10/86	0.00439
7	GO:0009653	anatomical structure morphogenesis	4/4	24/86	0.005
7	GO:0044092	negative regulation of molecular function	3/4	12/86	0.0079
7	GO:0060548	negative regulation of cell death	3/4	13/86	0.0102
7	GO:0005654	nucleoplasm	3/4	13/86	0.0102
7	GO:0043066	negative regulation of apoptotic process	3/4	13/86	0.0102
7	GO:0043069	negative regulation of programmed cell death	3/4	13/86	0.0102
7	GO:2000026	regulation of multicellular organismal development	4/4	29/86	0.0112
7	GO:0007166	cell surface receptor signaling pathway	3/4	14/86	0.0128
7	GO:0048729	tissue morphogenesis	3/4	14/86	0.0128
7	GO:0043067	regulation of programmed cell death	3/4	14/86	0.0128
7	GO:0042981	regulation of apoptotic process	3/4	14/86	0.0128
7	GO:0045595	regulation of cell differentiation	4/4	30/86	0.0129
7	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	4/4	30/86	0.0129
7	GO:0008284	positive regulation of cell proliferation	3/4	15/86	0.0159

7	GO:0010941	regulation of cell death	3/4	15/86	0.0159
7	GO:0050793	regulation of developmental process	4/4	32/86	0.0169
7	GO:0051239	regulation of multicellular organismal process	4/4	33/86	0.0193
7	GO:0071840	cellular component organization or biogenesis	3/4	16/86	0.0193
7	GO:0016043	cellular component organization	3/4	16/86	0.0193
7	GO:0051241	negative regulation of multicellular organismal process	3/4	17/86	0.0232
7	GO:0009966	regulation of signal transduction	3/4	17/86	0.0232
7	GO:0045596	negative regulation of cell differentiation	3/4	17/86	0.0232
7	GO:0010646	regulation of cell communication	3/4	18/86	0.0276
7	GO:0051093	negative regulation of developmental process	3/4	18/86	0.0276
7	GO:0048583	regulation of response to stimulus	3/4	18/86	0.0276
7	GO:0023051	regulation of signaling	3/4	18/86	0.0276
7	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	4/4	37/86	0.0311
7	GO:1903507	negative regulation of nucleic acid-templated transcription	4/4	37/86	0.0311
7	GO:1902679	negative regulation of RNA biosynthetic process	4/4	37/86	0.0311
7	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	4/4	37/86	0.0311
7	GO:0009890	negative regulation of biosynthetic process	4/4	37/86	0.0311
7	GO:0051172	negative regulation of nitrogen compound metabolic process	4/4	37/86	0.0311
7	GO:0045892	negative regulation of transcription, DNA-templated	4/4	37/86	0.0311
7	GO:0031327	negative regulation of cellular biosynthetic process	4/4	37/86	0.0311
7	GO:0010558	negative regulation of macromolecule biosynthetic process	4/4	37/86	0.0311
7	GO:0051253	negative regulation of RNA metabolic process	4/4	37/86	0.0311
7	GO:0044428	nuclear part	3/4	19/86	0.0324
7	GO:0010605	negative regulation of macromolecule metabolic process	4/4	38/86	0.0348
7	GO:0010629	negative regulation of gene expression	4/4	38/86	0.0348
7	GO:0031324	negative regulation of cellular metabolic process	4/4	38/86	0.0348
7	GO:0009892	negative regulation of metabolic process	4/4	39/86	0.0387
7	GO:0051094	positive regulation of developmental process	3/4	21/86	0.0435
7	GO:0044422	organelle part	3/4	22/86	0.0499
7	GO:0044446	intracellular organelle part	3/4	22/86	0.0499

7	GO:0051240	positive regulation of multicellular organismal process	3/4	22/86	0.0499
8	GO:1901362	organic cyclic compound biosynthetic process	0/3	66/86	0.0111
8	GO:1901360	organic cyclic compound metabolic process	0/3	66/86	0.0111
8	GO:0006139	nucleobase-containing compound metabolic process	0/3	66/86	0.0111
8	GO:0019438	aromatic compound biosynthetic process	0/3	66/86	0.0111
8	GO:1901576	organic substance biosynthetic process	0/3	66/86	0.0111
8	GO:0043170	macromolecule metabolic process	0/3	66/86	0.0111
8	GO:0018130	heterocycle biosynthetic process	0/3	66/86	0.0111
8	GO:0044271	cellular nitrogen compound biosynthetic process	0/3	66/86	0.0111
8	GO:0071704	organic substance metabolic process	0/3	66/86	0.0111
8	GO:0008152	metabolic process	0/3	66/86	0.0111
8	GO:0046483	heterocycle metabolic process	0/3	66/86	0.0111
8	GO:0044237	cellular metabolic process	0/3	66/86	0.0111
8	GO:0034654	nucleobase-containing compound biosynthetic process	0/3	66/86	0.0111
8	GO:0006725	cellular aromatic compound metabolic process	0/3	66/86	0.0111
8	GO:0009058	biosynthetic process	0/3	66/86	0.0111
8	GO:0009059	macromolecule biosynthetic process	0/3	66/86	0.0111
8	GO:0034641	cellular nitrogen compound metabolic process	0/3	66/86	0.0111
8	GO:0044238	primary metabolic process	0/3	66/86	0.0111
8	GO:0090304	nucleic acid metabolic process	0/3	66/86	0.0111
8	GO:0016070	RNA metabolic process	0/3	66/86	0.0111
8	GO:0044260	cellular macromolecule metabolic process	0/3	66/86	0.0111
8	GO:0032774	RNA biosynthetic process	0/3	66/86	0.0111
8	GO:0044249	cellular biosynthetic process	0/3	66/86	0.0111
8	GO:0006807	nitrogen compound metabolic process	0/3	66/86	0.0111
8	GO:0006351	transcription, DNA-templated	0/3	65/86	0.013
8	GO:0097659	nucleic acid-templated transcription	0/3	65/86	0.013
8	GO:0034645	cellular macromolecule biosynthetic process	0/3	65/86	0.013
9	GO:0006357	regulation of transcription from RNA polymerase II promoter	0/3	59/86	0.0286
9	GO:0001071	nucleic acid binding transcription factor activity	0/3	59/86	0.0286
9	GO:0003700	transcription factor activity, sequence-specific DNA binding	0/3	59/86	0.0286
12	GO:0030154	cell differentiation	5/5	41/86	0.0215
12	GO:0042127	regulation of cell proliferation	4/5	25/86	0.0237
12	GO:0048513	organ development	4/5	28/86	0.0369
12	GO:0045595	regulation of cell differentiation	4/5	30/86	0.0482

15	GO:0005667	transcription factor complex	3/4	22/86	0.0499
17	GO:0010001	glial cell differentiation	3/7	4/86	0.00132
17	GO:0048663	neuron fate commitment	3/7	5/86	0.00318
17	GO:0045665	negative regulation of neuron differentiation	3/7	6/86	0.00612
17	GO:0048731	system development	5/7	21/86	0.00856
17	GO:0045165	cell fate commitment	4/7	13/86	0.00893
17	GO:0050768	negative regulation of neurogenesis	3/7	8/86	0.0159
17	GO:0010721	negative regulation of cell development	3/7	8/86	0.0159
17	GO:0051961	negative regulation of nervous system development	3/7	9/86	0.0229
17	GO:0009605	response to external stimulus	3/7	9/86	0.0229
17	GO:0007399	nervous system development	3/7	9/86	0.0229
19	GO:0032502	developmental process	0/3	62/86	0.0198
19	GO:0044767	single-organism developmental process	0/3	58/86	0.032
19	GO:0044763	single-organism cellular process	0/3	57/86	0.0357
26	GO:0042802	identical protein binding	4/6	18/86	0.0161
26	GO:0042803	protein homodimerization activity	4/6	18/86	0.0161
26	GO:0048519	negative regulation of biological process	6/6	46/86	0.0281
26	GO:0070887	cellular response to chemical stimulus	3/6	12/86	0.0333
26	GO:0000989	transcription factor activity, transcription factor binding	4/6	24/86	0.0486
30	GO:0048513	organ development	5/6	28/86	0.0129
